IEEE Open Journal of Signal Processing

Received 28 September 2022; revised 29 December 2022; accepted 30 December 2022. Date of publication 23 January 2023; date of current version 10 March 2023. The review of this article was arranged by Associate Editor Arsenia Chorti.

Digital Object Identifier 10.1109/OJSP.2023.3238914

Multi-Exposure Image Compression Considering Rate-Distortion Optimization in Rendered High Dynamic Range Image

JUI-CHIU CHIANG [1] (Senior Member, IEEE), HUNG-YEN SHANG [2], AND JI-JIN QIU [1]

[1]Electrical Engineering Department, Center for Innovative Research on Aging Society, Advanced Institute of Manufacturing with High-Tech Innovation, National Chung Cheng University, Chia-Yi 621301, Taiwan
[2]Cubtex, Hsin-Chu 302082, Taiwan

CORRESPONDING AUTHOR: JUI-CHIU CHIANG (e-mail: rachel@ccu.edu.tw)

This work was supported in part by the National Science and Technology Council, Taiwan under Grant 107-2221-E-194-040, and in part by the Center for Innovative Research on Aging Society, Advanced Institute of Manufacturing with High-Tech Innovations from the Featured Areas Research Center Program through Higher Education Sprout Project by the Ministry of Education in Taiwan.

ABSTRACT High-dynamic range (HDR) images,which provide realistic visual perception, have attracted considerable attention in various applications. A simple method for obtaining HDR images is to fuse multiple low-dynamic range (LDR) images captured under varying exposures. We proposed a method of efficiently encoding multi-exposure images using multi-view High Efficiency Video Coding (MV-HEVC) along with intensity mapping function (IMF) to create high-quality HDR images. The inter-view prediction in MV-HEVC eliminates redundancy between multi-exposure images. To achieve high-efficiency MV-HEVC-based multi-exposure image coding, two modifications to the rate-distortion optimization (RDO) process performed during encoding were proposed. First, the distortion in the rendered HDR image is used to replace the original distortion, which depends on the LDR image. Second, to balance distortion and rate in RDO, a new Lagrange multiplier is derived from the derivative of the new distortion with respect to the rate. A model of the distortion in the HDR domain was constructed and expressed as a function of the distortion in the LDR domain. The proposed scheme utilizes the conventional 8-bit codec to compress multi-exposure LDR images and generate an HDR image in the decoder. The experimental results indicate that the proposed technique outperforms HDR image coding schemes in which HEVC Range Extension and JPEG-XT coding standards are used.

INDEX TERMS High dynamic range, multi-exposure image, multi-exposure image compression, rate-distortion optimization (RDO).

I. INTRODUCTION

Radiation in the real world spans several orders of magnitude. The conventional camera has limited capability to fully capture and present the details of scenes with a broad range of luminance. The images captured by the conventional camera are called low-dynamic range (LDR) images, which tend to lose the information in bright and dark areas. High-dynamic range (HDR) images have a wider dynamic range and contain rich information on the taken scenes. Although hardware devices that support HDR capture exist, they are usually too expensive and bulky to be widely used. The simple HDR image generation method using the conventional camera involves merging a series of LDR images captured under different exposure levels, which are called multi-exposure images, into an HDR image to create a more realistic image.

Numerous studies have proposed methods of obtaining a high-quality HDR images from multi-exposure images [1], [2], [3], [4]. Efficient compression is a crucial task in the production of HDR images. HDR images have considerably higher storage requirements than do LDR images because HDR images have a wider dynamic range. Because HDR images are stored in floating-point format [5], [6], [7], a preprocessing procedure that involves converting the floating-point format to the integer format must usually be performed

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

before HDR image coding under the current standards, most of which only support the integer format (JPEG-XT [8] supports both the integer and floating-point formats). The floating-point format is usually converted to the integer format by accounting for the characteristics of human vision to achieve superior coding performance and create a realistic image [9], [10], [11], [12]. Coding involving this conversion is called single-layer coding. A drawback of single-layer coding is the lack of backward compatibility. If only an LDR device is available, HDR images must be converted into LDR images through tone mapping [13], [14], [15]. To support backward compatibility, a two-layer coding method was proposed [8], [16], [17], [18], [19], [20], [21], [22], [23]. In this method, an LDR image can be reconstructed from a base layer alone, whereas an HDR image can be generated from a base and enhancement layer in the decoder. However, in this coding method, the LDR image is a tone-mapped version of the input HDR image, and tone mapping and inverse tone mapping affecting coding efficiency differently. Therefore, several studies have explored tone mapping and inverse tone mapping algorithms to increase the efficiency of HDR image coding [17], [19], [20], [21].

Most image and video coding standards only support content in the integer format, and their encoding procedures are optimized for LDR images. Floating-value HDR image compression based on the conventional coding architecture is not guaranteed to have optimal coding efficiency.

We proposed an HDR image coding technique. In the proposed technique, instead of encoding the HDR image, which is generated from stacked LDR images, multi-exposure images are compressed, and HDR images are generated in the decoder. Because multi-exposure images present the same scene under different exposure levels, they are highly correlated and high coding efficiency is expected if the redundancy among multi-exposure images is well exploited. To achieve this goal, multi-view architecture is adopted in this study where the inter-view prediction is used to eliminate the redundancy during encoding.

Multi-exposure images of a scene are captured to generate an HDR image. The main concern during the encoding of multi-exposure images is the trade-off between the bitrate and the distortion of the generated HDR image. A high coding efficiency for each LDR image in the multi-exposure stack may not lead to a high-quality HDR image. The HDR coding efficiency can be increased if the HDR distortion is considered during the coding of multi-exposure images. This is the first work to consider HDR distortion during the encoding of multi-exposure images. The distortion of an HDR image is obtained using the inverse camera response function (ICRF). Distortion in the HDR domain can be computed after the multi-exposure images are reconstructed. In the proposed method, the distortion term is changed from the original LDR image distortion to the HDR image distortion during rate-distortion optimization (RDO). To balance distortion and rate, a new Lagrange multiplier is used. The chain rule is used to split the derivative of the distortion in the HDR image with respect to the rate into two terms, and the distortion of an HDR image is expressed in terms of the distortion of the corresponding LDR image.

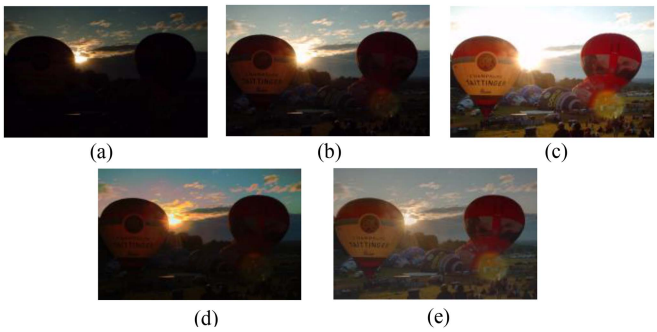


FIGURE 1. Multi-exposure images: (a) under-exposed image, (b) well-exposed image, (c) over-exposed image, (d) tone-mapped image [13], and (e) MEF image [24].

TABLE 1. Table of Notation

symbol	description
V	the original view V
$\hat{V}$	the reconstructed view V
D_V	CTU distortion of view V in the LDR domain
$D_{V,HDR}$	CTU distortion of view V in the HDR domain
$D_V^{non-base}$	CTU distortion of nonbase view V
R_V	bits to encode the current CTU of view V
$\lambda_{V,HDR}$	Lagrange multiplier for nonbase view V
$\varphi(\cdot)$	intensity mapping function
$f(\cdot)$	camera response function (CRF)
$L_{i,j}$	pixel j in an LDR image with exposure i
H	irradiance
N	image capture noise
$f^{-1}(\cdot)$	inverse camera response function (ICRF)
Z	output of ICRF
w	weight, reflecting the influence of N
D_{Z_i}	CTU distortion in terms of Z value of view V_i
$DR_{i\to m}$	ratio between D_{Z_i} and D_{Z_m}

In addition, using the proposed technique, a high-quality LDR image can also be obtained by fusing reconstructed multi-exposure LDR images. Multi-exposure fused (MEF) LDR images [24], [25], [26], [27], which are LDR images with enhanced contrast, exhibit higher visual quality than do tone-mapped versions of reconstructed HDR images, particularly in oversaturated and undersaturated regions (Fig. 1).

The main contributions of this paper are as follows:

1) A multi-view-based multi-exposure image coding scheme was proposed. The scheme is based on the conventional 8-bit codec and achieves coding performance superior to that of conventional coding methods, in which HDR images generated from the multi-exposure images are encoded using the High Efficiency Video Coding (HEVC) and JPEG-XT coding standards.

2) An RDO modification was introduced for the proposed multi-exposure image coding framework. In this modification, the distortion of an HDR image is used, rather than the distortion of an LDR image.
3) A new Lagrange multiplier was derived to balance the rate and the modified distortion.

The rest of this paper is organized as follows. Section II presents a brief review of studies related to RDO in image and video compression as well as HDR image and video coding. Section III describes the proposed scheme. Section IV presents and discusses the experimental results. Finally, Section V presents the conclusions.

II. RELATED WORKS

A. RATE-DISTORTION OPTIMIZATION

For image and video coding, RDO [28] is used to determine optimal coding mode θ^*, which balances cost, namely the rate, with performance, namely the distortion. In RDO, optimal coding mode θ^* is identified by minimizing cost J as follows:

$$\theta^* = \min_{\theta} J(\theta; \lambda) = D(\theta) + \lambda \cdot R(\theta) \tag{1}$$

where θ is the candidate coding mode, which may involve the use of parameters such as the reference frame, coding block size, and quantization parameter (QP); D is the distortion of the current block, which is usually equal to the sum of the squared difference (SSD) between the reconstructed block and the uncompressed block; R represents the bits required for encoding the current block; and λ is a Lagrange multiplier used to balance D with R. In [28], experiments were conducted to identify the optimal Lagrange multiplier. First, several Lagrange multipliers were selected, and the quantization setting resulting in the minimum rate–distortion (RD) cost was recorded for each Lagrange multiplier. After the relative occurrence of the quantization settings under different Lagrange parameters was identified, the relationship between λ and the quantization settings was regressively determined using a formula. Another method used in [28] to determine the optimal Lagrange multiplier involved directly deriving the mathematical relationship between the rate and distortion. Under the assumption that the distortion D and the rate R were differentiable everywhere, the optimal Lagrange multiplier was identified by setting the derivative of cost J in (1) to 0.

$$\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0 \tag{2}$$

On the basis of (2), the following equation can be obtained: $\lambda = -\frac{dD}{dR}$. Under high-rate assumption, the source probability distribution can be approximated as a constant within each quantization interval, and the distortion can be expressed as follows:

$$D(Q) = \frac{(2Q)^2}{12} = \frac{Q^2}{3}, \tag{3}$$

where $2Q$ is the quantization step size. Under the assumption of a high rate, rate R is related to distortion D as follows:

$$R(D) = \alpha \log_2\left(\frac{\beta}{D}\right) \tag{4}$$

where α and β are constants. Parameter λ can be described as a function of Q as follows:

$$\lambda = -\frac{dD}{dR} = \gamma \cdot Q^2, \tag{5}$$

where γ is a constant. The relationship expressed in (5) is similar to that obtained in the experiments conducted in [28] on the relative occurrence of the quantization setting for a given Lagrange multiplier. Eq. (5) has been incorporated into various coding standards, such as H.264 [29], HEVC [30], and Versatile Video Coding (VVC) [31]. In H.264 and HEVC, (5) is rewritten as follows:

$$\lambda = -\frac{dD}{dR} = \gamma \cdot Q^2 = s_f \cdot 2^{(QP-12)/3} \tag{6}$$

where s_f is a factor related to the coding frame property. The λ values of the I frame, P frame (predicted frame), and B frame (bidirectional predicted frame) used in the H.264, HEVC, and VVC formats differ. In HEVC, λ values are defined as follows:

$$\begin{cases} \lambda_I = (1 - \max(0,\ \min(0.5,\ 0.05N_B))) \times 0.57 \cdot 2^{\frac{QP-12}{3}} \\ \lambda_P = u \times 2^{\frac{QP-12}{3}} \\ \lambda_B = u \times \max(2,\ \min(4,\ (QP-12)/6)) \times 2^{\frac{QP-12}{3}} \end{cases} \tag{7}$$

where u is a scaling factor that can be differently configured for frames at different temporal levels in a group of pictures (GOP) and N_B is the number of consecutive B frames in a GOP.

Several researchers have modified the Lagrange multiplier to increase coding efficiency. In [32], the optimal Lagrange multiplier for the B frame was predicted using the ratio between the average distortions of the P and B frames. A joint optimization method for identifying a suitable coding mode and Lagrange multiplier for interframe coding was presented in [33]. To accurately reflect the perceived image quality in RDO, numerous studies have incorporated the properties of the human visual system into image coding [34], [35], [36], [37]. Structural similarity index measure (SSIM)-based RDO methods were proposed in [34] and [35]. The methods used in these studies are based on a similar idea but different definitions of cost function J ($(1 - SSIM) + \lambda R$ in [34] and $(1/SSIM) + \lambda R$ in [35]). The process of determining optimal Lagrange multiplier λ also differed between the methods. In [34], a λ value was selected such that the rate remained the same regardless of whether the mean square error (MSE) or 1/SSIM was used as the distortion metric. In [35], λ was obtained by setting the derivative of J with respect to R to 0. Thus, the following equation was obtained in [35]:

$$\frac{dJ}{dR} = \frac{dSSIM}{dR} + \lambda = 0, \lambda = -\frac{dSSIM}{dR} = -\frac{\frac{dSSIM}{dq}}{\frac{dR}{dq}} \tag{8}$$

In [35], the SSIM and rate models were constructed in terms of q, which represents the quantization step for deriving λ. A type of perceptual distortion was discussed in [38]. In [38], the relationship between an objective metric, namely the MSE, and the results of a subjective test was determined. The perceived distortion was directly proportional to the MSE under scaling factor η. The original λ value used in HEVC was modified by multiplying it by η.

B. HDR IMAGE AND VIDEO CODING

HDR image and video coding methods can be categorized into two groups: single- and two-layer coding methods. Single-layer coding has a higher efficiency than does two-layer coding but does not support backward compatibility. Two-layer coding offers backward compatibility for legacy standard dynamic range (SDR) devices but requires a higher bit consumption than does single-layer coding. The following section reviews studies on these methods.

1) SINGLE-LAYER HDR IMAGE CODING

Because most image and video coding standards, such as H.264, HEVC, and VVC, allow only input in the integer format, a preprocessing procedure must usually be employed in single-layer coding to convert a floating-point HDR image into the integer format by using specific transfer functions.

In [39], HDR images in the OpenEXR format were encoded using the H.264/AVC standard after being transformed into an adaptive LogLuv format. Two adaptive parameters were used to scale the range of Luv values and to utilize the dynamic range by referring to the maximum and minimum brightness of the HDR. In [40], the luminance masking effect was considered, and the size of the quantization step was determined for each transform unit on the basis of intensity. Perceptual quantization has been adopted in several studies [10], [11] to convert the floating-point format into the integer format. An adaptive perceptual quantizer (PQ) was developed in [10], in which the codeword was designed by considering the minimum detectable contrast and luminance distribution of the HDR content. Numerous researchers have modified the RDO process to enhance the visual quality of LDR images [32], [33], [34], [35], [38]. Perceptual uniform (PU) encoding has been used to modify the distortion during the RDO process for increasing the efficiency of HDR image coding [12], [41]. In [42], large-scale testing related to single-layer HDR video coding was performed, with performance comparisons being conducted for four transfer functions, three color encoding methods, and two coding schemes.

AVC, HEVC and VVC [31] support high-bit depth video coding and require floating-point HDR sequences to be converted into the integer format before encoding. Two types of transfer functions are usually used: the PQ and hybrid log–gamma (HLG) transfer functions [43], [44]. In PQ and HLG systems, three transfer functions are defined: the electro-optical transfer function (EOTF), which defines the mapping between nonlinear color signal values and the display light; opto-optical transfer function, which defines the mapping between the relative scene light and the display light; and opto-electro transfer function (OETF), which defines the mapping between the relative scene light and nonlinear color values. The OETF is used in encoders, and the EOTF is used in decoders for display.

To evaluate the performance of HDR image and video coding in VVC, four objective metrics are used [45]: the peak signal-to-noise ratio (PSNR), the weighted PSNR (WPSNR), deltaE100, and PSNR-L100. A weighted MSE is used to calculate the WPSNR, with the weight depending on the luma value corresponding to the pixel. The deltaE100 distortion indicates the distance in the Lab color space, and PSNR-L100 represents the distortion in the lightness domain of the CIELab color space.

2) TWO-LAYER HDR IMAGE CODING

The two-layer coding scheme supports backward compatibility by providing a base layer and an enhancement layer. The base layer contains tone-mapped LDR images generated from input HDR images, and this layer is usually encoded using a legacy 8-bit codec. The enhancement layer contains ratio- and residual-based information. The ratio-based enhancement layer encodes the ratio between the original HDR and the reconstructed base layer. The residual-based enhancement layer encodes the residual between the original HDR image and the predicted HDR image, which is the inverse tone-mapped version of the decoded LDR image. Tone mapping and inverse tone mapping are essential steps in two-layer image coding. JPEG-XT [8] is a two-layer coding scheme that supports integer and floating-point HDR image coding. The JPEG-XT scheme contains three profiles: Profile A, Profile B, and Profile C. These three profiles have the same base layer, which is encoded using a legacy JPEG encoder. The enhancement layers of Profile A and Profile B provide ratio-based information but different computation of ratio images. The enhancement layer in Profile C provides residual-based information.

Studies have explored two-layer HDR coding more than single-layer HDR coding. A JPEG backward-compatible coding scheme was proposed in [46]. In this scheme, the enhancement layer contains residual-based information, and HDR prediction is performed in the transform domain. The authors of [17] explored inverse tone mapping for two-layer HDR image coding. The relationship between the base layer and the HDR image was used to estimate the inverse tone mapping curve. To compare the coding efficiencies of various decoding tools, a large-scale objective quality test was conducted in [47]. An objective evaluation of HDR-VDP-2 [48] was conducted in [49] by considering various parameters, including the adopted tone mapping algorithm and the quality of the base layer and residual layer. An interlayer color prediction method was proposed in [50]. In this method, the chroma of a decoded LDR image is used to predict the chroma of the corresponding HDR image. Learning-based image and video coding techniques have attracted considerable research

attention [51], [52]. In [53], HDR image compression was conducted using a convolutional autoencoder; the base layer was encoded using a JPEG encoder, and the enhancement layer was encoded using a neural network.

A two-layer coding scheme for HDR videos was first reported in [16]. In the study, the residual-based architecture was adopted, and the enhancement layer was filtered to eliminate the psychovisual redundancy before encoding. Several two-layer coding techniques involve the use of tone mapping to increase HDR video coding efficiency [19], [20], [21], [22]. In [19], a piecewise tone mapping model was proposed. This model derives the optimal parameters to minimize distortion in the logarithmic domain for each segment. In [20], the distortion in the PU domain was considered instead. Tone mapping was conducted in [21] and [22] by optimizing the RD performance and satisfying certain constraints. An LDR quality constraint was used in [21], and the brightness difference between consecutive LDR frames was considered in [22]. In [23], a template-based interlayer prediction method was developed to decrease the signaling overhead for the parameters of inverse tone mapping.

In general, HDR images are generated using multi-exposure LDR images. In this study, we attempted to achieve high-efficiency multi-exposure LDR image encoding to generate a high-quality HDR image in the decoder. In our previous study [54], multi-exposure LDR images were coded using a multi-view coding scheme in which each LDR image is treated as a view and the redundancy between views is used after intensity conversion between the views. The main task in [54] was to redefine the distortion in the RDO process. In the RDO process adopted in [54], both the distortion in the current block and MEF block was used to define the distortion. However, an MEF image is a contrast-enhanced LDR image, and the distortion in an MEF image does not completely reflect the quality of a reconstructed HDR image.

In this study, to accurately reflect the quality of the generated HDR image using compressed multi-exposure images, the original LDR distortion in MV-HEVC was replaced with HDR distortion for nonbase view encoding in MV-HEVC-based multi-exposure image coding. To balance distortion and rate during RDO, a new Lagrange multiplier was derived from the derivative of the HDR distortion with respect to the rate. To obtain this derivative, the distortion in the HDR image is expressed as a function of the distortion in the LDR image. The experimental results indicate the superiority of the proposed technique to those proposed in other studies in terms of HDR-VDP2.

III. PROPOSED FRAMEWORK

In the proposed method, multi-exposure images are encoded instead of HDR images being generated because multi-exposure images contain high redundancy. The efficiency of multi-exposure image coding depends on how well the correlation among the images is explored. As in our previous study [54], this study used an MV-HEVC [55] scheme, and each LDR image was treated as a view. The use of this conventional

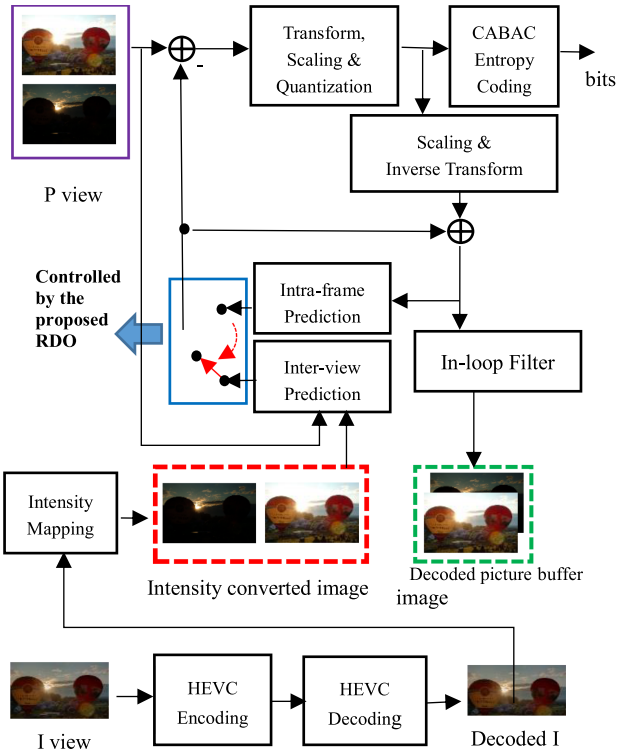


FIGURE 2. Framework of proposed coding scheme.

8-bit codec for HDR image coding is practical because it does not require conversion from the floating-point format to the integer format where the PQ operation is employed in general HDR image coding.

In MV-HEVC, one view serves as the base view (called the "I view"), whereas the remaining views are nonbase views (called "P views"). The I view is encoded through HEVC, and the P views are encoded through intra-frame prediction and inter-view prediction. During inter-view prediction, the reconstructed I view is used as a reference frame. Because the brightness of multi-exposure images differs, the inter-view redundancy cannot be suitably exploited, and the coding efficiency does not surpass that of simulcast coding. To increase the accuracy of inter-view prediction, the brightness variation among views should be minimized as small as possible. Therefore, the exposure of the reconstructed I view is changed before the encoding of nonbase views.

Fig. 2 displays the framework of the proposed coding scheme, where images with three types of exposure, namely a well-exposed image, an under-exposed image, and an over-exposed image, are considered. The well-exposed image is considered the base view (I view), whereas the other two images are considered nonbase views (P views). The encoding policy of the nonbase view changes, whereas that of the base view does not. For the base view, only intra-frame prediction is allowed, followed by transform coding, quantization and entropy coding. For the nonbase views, the mode decision between the intra-frame prediction and inter-view prediction depends on the new RDO principle. To better exploit the

redundancy among views, the inter-view prediction is performed after the intensity of the decoded I frame is converted through intensity mapping. It means the reconstructed base view changes its intensity and becomes similar to that of the current nonbase view. In the RDO process, the distortion and Lagrange multiplier are modified after taking the HDR distortion into consideration. The proposed framework can be applied to multi-exposure images with various exposures.

A. COMPUTATION OF INTENSITY MAPPING FUNCTION AND ESTIMATION OF ICRF

Before the encoding of multi-exposure images, the intensity mapping function (IMF) between the base view and a nonbase view is determined and the ICRF is estimated using multi-exposure LDR images. IMF is used to enable superior inter-view prediction while ICRF is used to generate an HDR image from multi-exposure LDR images. In the proposed framework, the ICRF is also used to compute the distortion in the generated HDR image during the encoding of the nonbase views. Additional details related to this function are provided in the following sections.

The base view is encoded first, followed by the nonbase views. Before the encoding of the nonbase views, intensity conversion is performed for the reconstructed base view to ensure its brightness similarly to that of the nonbase views. This step is performed to increase the encoding efficiency of the nonbase views during inter-view prediction. Cumulative histogram matching is used to compute the IMF (φ) [4] for the current nonbase view and base view as follows:

$$T_1(\varphi(d)) = T_2(d) \rightarrow \varphi(d) = T_2^{-1}(T_1(d)), \tag{9}$$

where d is the pixel intensity and T_1 and T_2 represent the cumulative histograms for the base view and current nonbase view, respectively.

B. PROPOSED MULTI-EXPOSURE IMAGE CODING SCHEME

Multi-exposure images are encoded using MV-HEVC in the proposed scheme. In the case of three images with different exposure levels, namely a well-exposed image, an under-exposed image, and an over-exposed image, the well-exposed image is treated as the base view and encoded as the I view, whereas the under-exposed and over-exposed images are nonbase views that are encoded as P views. A reconstructed base view is used as a reference during inter-view prediction. The well-exposed image is encoded first, followed by the under-exposed and over-exposed images. The LDR images are referred to as "View 0" (V_0), "View 1" (V_1), and "View 2" (V_2) on the basis of the coding order in the proposed architecture. In [54], a change in coding order between the under-exposed and over-exposed images did not affect coding performance because nonbase view uses only the base view as reference.

Multi-exposure images are used to generate HDR images. If each view is encoded through a conventional RDO process, the coding efficiency is optimized for each view but may not be optimized for the generated HDR image. Because the ultimate goal is to reconstruct a high-quality HDR image, the distortion of the generated HDR image should be considered and not the distortion of the LDR image. Bitrate waste occurs if pixels with a low contribution to HDR imaging are subjected to high-quality encoding. In the proposed scheme, to encode nonbase views, a modified distortion is used in the RDO process. The distortion in the HDR domain is considered in the RDO process of the proposed scheme. To obtain this distortion, the current view and previously encoded view are required. The method for computing the HDR distortion is detailed in the following text.

1) DISTORTION MODIFICATION FOR NONBASE VIEWS

The RDO process is briefly introduced in Section II-A. RDO is conducted to determine the optimal coding mode for a certain distortion, rate, and Lagrange multiplier. Distortion D_V of the current coding tree unit (CTU) in HEVC is expressed as follows:

$$D_V = \sum_{(x,y)\in B} \left(V(x,y) - \hat{V}(x,y)\right)^2, \tag{10}$$

where V and $\hat{V}$ are the original and the reconstructed images, respectively, and B represents the current CTU.

During the encoding of the base view, other views are not encoded or referred to. Therefore, the original RDO process is employed directly. However, for each CTU of a nonbase view, the definitions of distortion in the RDO process $D_V^{non-base}$ and cost J are modified as follows:

$$D_V^{non-base} = D_{V,\ HDR} \tag{11}$$

$$J = D_{V,\ HDR} + \lambda_{V,\ HDR} \cdot R_V, \tag{12}$$

where $D_V^{non-base}$ is the CTU distortion in nonbase view V and is obtained by computing the distortion in the generated HDR CTU, which is represented by $D_{V,\ HDR}$. Parameter $\lambda_{V,\ HDR}$ represents the modified Lagrange multiplier for nonbase view V. To determine $D_{V,HDR}$, the difference between the generated HDR CTU in the encoder and the reconstructed HDR CTU in the decoder is computed. The following section introduces the Robertson camera response function (CRF) model [56], which is used to generate an HDR image from multi-exposure images. Subsequently, the derivation of Lagrange multiplier $\lambda_{V,HDR}$ for the RDO process during the encoding of the nonbase view is described.

2) ROBERTSON CRF MODEL

Given N images of a static scene with different exposure times (t_i, $i = 1 \ldots N$), LDR image L can be obtained through mapping based on the following function:

$$L_{i,j} = f\left(t_i H_j + N_{i,j}\right), \tag{13}$$

where $f(\cdot)$ denotes the CRF; j is the pixel index; H is the irradiance; and $N_{i,j}$ is the additive noise term, modeled as a Gaussian distribution with a mean of 0 and a variance of σ_{ij}^2.

For HDR imaging, irradiance H is estimated for the observation L. The ICRF is represented by $f^{-1}(\cdot)$ and can be

expressed as follows:

$$f^{-1}\left(L_{i,j}\right) = t_i H_j + N_{i,j} = Z_{i,j}, \tag{14}$$

where $Z_{i,j}$ can be modeled as independent Gaussian variables. The joint probability function can be expressed as follows:

$$\mathrm{P}\left(\mathbf{Z}\right) \propto \exp\left(-\frac{1}{2}\sum_{i,j} w_{i,j}\left(Z_{i,j} - t_i H_j\right)^2\right), \tag{15}$$

where $w_{i,j} = 1/\sigma_{i,j}^2$. Parameter $w_{i,j}$ is modeled using a Gaussian-like function as follows:

$$w_{i,j} = w\left(L_{i,j}\right) = \exp\left(-4 \cdot \left(L_{i,j} - 127.5\right)^2 / 127.5^2\right),$$
$$\text{and } w\left(0\right) = w\left(255\right) = 0, \quad w\left(127.5\right) = 1 \tag{16}$$

Because Z and H are unknown, the objective function is

$$O\left(H,\ Z\right) = \sum_{i,j} w_{i,j}\left(Z_{i,j} - t_i H_j\right)^2 \tag{17}$$

A form of Gauss–Seidel relaxation is used to iteratively determine the solution, and HDR image H in the kth iteration can be computed as follows:

$$\hat{H}_j^k = \frac{\sum_i w_{i,j} t_i \hat{Z}_{i,j}^k}{\sum_i w_{i,j} t_i^2}, \tag{18}$$

where $\hat{Z}^k$ is the Z value estimated in the kth iteration and is computed using $\hat{H}^{k-1}$. The final estimate of the HDR image is obtained when the convergence condition is satisfied. The HDR CTU distortion can be obtained after the HDR values in the encoder and decoder have been computed using (18).

3) MODIFICATION OF LAGRANGE MULTIPLIER

Because the definition of distortion is modified, the value of the Lagrange multiplier must be changed accordingly to maintain a suitable balance between the new distortion and the rate. As described in Section II-A, an optimal λ value for the RDO process can be determined from the derivative of the distortion with respect to the rate; thus, the following equation is obtained:

$$\lambda = -\frac{dD}{dR} \tag{19}$$

If the modified distortion is expressed as a function of the rate, the Lagrange multiplier can be easily computed. On the basis of this principle, the optimal value for $\lambda_{V,HDR}$ in (12) is determined through a chain rule derivative as follows:

$$\lambda_{V,HDR} = -\frac{dD_{V,HDR}}{dR_V} = \frac{dD_{V,HDR}}{dD_V} \cdot \left(-\frac{dD_V}{dR_V}\right)$$
$$= \frac{dD_{V,HDR}}{dD_V} \cdot \lambda = \lambda_{V,s\_HDR} \cdot \lambda, \tag{20}$$

where D_V is the distortion of the current CTU in the LDR image, λ_{V,s_HDR} $(= dD_{V,HDR}/dD_V)$ is the derivative of $D_{V,HDR}$ with respect to D_V; and λ_{V,s_HDR} is a scaling factor applied to λ, which is the original Lagrange multiplier used in MV-HEVC. Parameter λ_{V,s_HDR} is determined after $D_{V,HDR}$ is modeled as a function of D_V.

For simplicity, we considered a two-view case in which views V_0 and V_1 have exposure times t_0 and t_1 , respectively. View V_0 is the base view, and the encoding procedure does not change for this view. To encode V_1, distortion in the HDR domain $D_{V_1,HDR}$ is considered in the RDO process. This parameter can be expressed as the SSD between original HDR value H_{V_1} and reconstructed HDR value $\hat{H}_{V_1}$, as presented in (21) shown at the bottom of the next page. The HDR values are computed using (18) when the convergence condition is satisfied.

In (21), L_0 and L_1 are the original LDR values of V_0 and V_1 , respectively, and $\hat{L}_0$ and $\hat{L}_1$ are the reconstructed LDR values of V_0 and V_1 , respectively. Besides,

$$C_j = \sum_{i=0,1} w_{i,j} t_i^2,\ \widehat{C_j} = \sum_{i=0,1} \hat{w}_{i,j} t_i^2$$
$$\widehat{C_j} \cong C_j \tag{22}$$

And b_0 and b_1 are the mean values of $\frac{w_{0,j}^2}{C_j^2}$ and $\frac{w_{1,j}^2}{C_j^2}$ in a CTU, respectively.

D_{Z_0} and D_{Z_1} are defined as:

$$D_{Z_0} = \sum_{j\in B}\left(f^{-1}\left(L_{0,j}\right) - f^{-1}\left(\hat{L}_{0,j}\right)\right)^2,$$
$$\text{and } D_{Z_1} = \sum_{j\in B}\left(f^{-1}\left(L_{1,j}\right) - f^{-1}\left(\hat{L}_{1,j}\right)\right)^2 \tag{23}$$

In (21), term $2\sum_{j\in B}\frac{1}{C_j^2} w_{0,j} w_{1,j} t_0 t_1 (f^{-1}(L_{0,j}) - f^{-1}(\hat{L}_{0,j}))$ $(f^{-1}(L_{1,j}) - f^{-1}(\hat{L}_{1,j}))$ is assumed to be 0 because the coding errors of V_0 and V_1 are independent of each other, as are terms $(f^{-1}(L_{0,j}) - f^{-1}(\hat{L}_{0,j}))$ and $(f^{-1}(L_{1,j}) - f^{-1}(\hat{L}_{1,j}))$.

The distortions for a given CTU in the LDR domain for V_0 and V_1 are represented by D_{V_0} and D_{V_1} , respectively, which are expressed as follows:

$$D_{V_0} = \sum_{j\in B}\left(L_{0,j} - \hat{L}_{0,j}\right)^2,$$
$$D_{V_1} = \sum_{j\in B}\left(L_{1,j} - \hat{L}_{1,j}\right)^2 \tag{24}$$

In (21), D_{V_1,s_HDR} is modeled in terms of D_{Z_1}. To derive the corresponding $\lambda_{V_1,HDR}$ $(= dD_{V_1,HDR}/dD_{V_1})$ value, (21) must be modeled in terms of D_{V_1}, which enables the derivative of the HDR distortion with respect to LDR distortion D_{V_1} to be obtained. It means that D_{Z_1} must be modeled as a function of D_{V_1}. To derive the relationship between D_{Z_1} and D_{V_1}, we need ICRF. In this paper, ICRF $(f^{-1}(\cdot))$ is assumed to be a second-order polynomial that can be expressed as follows:

$$Z = f^{-1}\left(L\right) = a_0 + a_1 L + a_2 L^2 \tag{25}$$

Under this assumption, D_{Z_1} can be rewritten as follows:

$$\begin{aligned}
D_{Z_1} &= \sum_{j\in B}\left(f^{-1}\left(L_{1,j}\right)-f^{-1}\left(\hat{L}_{1,j}\right)\right)^2 \\
&= \sum_{j\in B}\left(\left(a_0+a_1L_{1,j}+a_2L_{1,j}^2\right)\right. \\
&\quad \left.-\left(a_0+a_1\hat{L}_{1,j}+a_2\hat{L}_{1,j}^2\right)\right)^2 \\
&= \sum_{j\in B}\left(a_1\left(L_{1,j}-\hat{L}_{1,j}\right)+a_2\left(L_{1,j}^2-\hat{L}_{1,j}^2\right)\right)^2 \\
&= \sum_{j\in B}a_1^2\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \\
&\quad +\sum_{j\in B}a_2^2\left(L_{1,j}^2-\hat{L}_{1,j}^2\right)^2 \\
&\quad +\sum_{j\in B}2a_1a_2\left(L_{1,j}-\hat{L}_{1,j}\right)\left(L_{1,j}^2-\hat{L}_{1,j}^2\right) \\
&= a_1^2\sum_{j\in B}\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \\
&\quad +a_2^2\sum_{j\in B}\left(L_{1,j}+\hat{L}_{1,j}\right)^2\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \\
&\quad +2a_1a_2\sum_{j\in B}\left(L_{1,j}+\hat{L}_{1,j}\right)\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \qquad (26)
\end{aligned}$$

To express D_{Z_1} in terms of D_{V_1}, which is $\sum_{j\in B}\left(L_{1,j}-\hat{L}_{1,j}\right)^2$, as shown in (24), the second and third terms in (26) must be approximated to convert them into functions of D_{V_1}. The third term of (26) is approximated as follows:

$$\begin{aligned}
&2a_1a_2\sum_{j\in B}\left(L_{1,j}+\hat{L}_{1,j}\right)\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \\
&\quad\approx 2a_1a_2\left(L_{1,M}+\hat{L}_{1,M}\right)\sum_{j\in B}\left(L_{1,j}-\hat{L}_{1,j}\right)^2, \qquad (27)
\end{aligned}$$

where $L_{1,M}$ and $\hat{L}_{1,M}$ are the mean values of L_1 and $\hat{L}_1$ in a CTU of V_1, respectively. Similarly, the second term of (26) is approximated as follows:

$$\begin{aligned}
&a_2^2\sum_{j\in B}\left(L_{1,j}+\hat{L}_{1,j}\right)^2\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \\
&\quad\approx a_2^2\left(L_{1,M}+\hat{L}_{1,M}\right)^2\sum_{j\in B}\left(L_{1,j}-\hat{L}_{1,j}\right)^2 \qquad (28)
\end{aligned}$$

Thus, D_{Z_1} can be expressed in terms of D_{V_1} as follows:

$$D_{Z_1} = \left(a_1+a_2\left(L_{1,M}+\hat{L}_{1,M}\right)\right)^2 D_{V_1} \qquad (29)$$

Subsequently, (21) can be rewritten as follows:

$$\begin{aligned}
D_{V_1,HDR} &= \sum_{j\in B}\left(H_{V_1,j}-\hat{H}_{V_1,j}\right)^2 = \sum_{j\in B}\left(\frac{\sum_{i=0,1}w_{i,j}t_iZ_{i,j}}{\sum_{i=0,1}w_{i,j}t_i^2}-\frac{\sum_{i=0,1}\hat{w}_{i,j}t_i\hat{Z}_{i,j}}{\sum_{i=0,1}\hat{w}_{i,j}t_i^2}\right)^2 \\
&= \sum_{j\in B}\left(\frac{\sum_{i=0,1}w_{i,j}t_if^{-1}\left(L_{i,j}\right)}{C_j}-\frac{\sum_{i=0,1}\hat{w}_{i,j}t_if^{-1}\left(\hat{L}_{i,j}\right)}{\widehat{C}_j}\right)^2 \\
&\cong \sum_{j\in B}\left(\left[\frac{1}{C_j}\left(w_{0,j}t_0f^{-1}\left(L_{0,j}\right)+w_{1,j}t_1f^{-1}\left(L_{1,j}\right)\right)-\frac{1}{C_j}\left(\hat{w}_{0,j}t_0f^{-1}\left(\hat{L}_{0,j}\right)+\hat{w}_{1,j}t_1f^{-1}\left(\hat{L}_{1,j}\right)\right)\right]\right)^2 \\
&\cong \sum_{j\in B}\left(\left[\frac{w_{0,j}}{C_j}t_0\left(f^{-1}\left(L_{0,j}\right)-f^{-1}\left(\hat{y}_{0,j}\right)\right)+\frac{w_{1,j}}{C_j}t_1\left(f^{-1}\left(L_{1,j}\right)-f^{-1}\left(\hat{L}_{1,j}\right)\right)\right]\right)^2 \\
&= \sum_{j\in B}\left[\frac{w_{0,j}^2}{C_j^2}t_0^2\left(f^{-1}\left(L_{0,j}\right)-f^{-1}\left(\hat{L}_{0,j}\right)\right)^2+\frac{w_{1,j}^2}{C_j^2}t_1^2\left(f^{-1}\left(L_{1,j}\right)-f^{-1}\left(\hat{L}_{1,j}\right)\right)^2\right] \\
&\quad +2\sum_{j\in B}\frac{1}{C_j^2}w_{0,j}w_{1,j}t_0t_1\left(f^{-1}\left(L_{0,j}\right)-f^{-1}\left(\hat{L}_{0,j}\right)\right)\left(f^{-1}\left(L_{1,j}\right)-f^{-1}\left(\hat{L}_{1,j}\right)\right) \\
&\cong \sum_{j\in B}\left[\frac{w_{0,j}^2}{C_j^2}t_0^2\left(f^{-1}\left(L_{0,j}\right)-f^{-1}\left(\hat{L}_{0,j}\right)\right)^2+\frac{w_{1,j}^2}{C_j^2}t_1^2\left(f^{-1}\left(L_{1,j}\right)-f^{-1}\left(\hat{L}_{1,j}\right)\right)^2\right] \\
&\cong \left(b_0t_0^2D_{Z_0}+b_1t_1^2D_{Z_1}\right) = \left(b_0t_0^2\frac{D_{Z_0}}{D_{Z_1}}+b_1t_1^2\right)D_{Z_1} \qquad (21)
\end{aligned}$$

$$= \left(b_0 t_0^2 \frac{D_{Z_0}}{D_{Z_1}} + b_1 t_1^2\right) \left(a_1 + a_2 \left(L_{1,M} + \hat{L}_{1,M}\right)\right)^2 D_{V_1} \tag{30}$$

Then, parameter λ_{V_1,s_HDR} can be derived as follows:

$$\lambda_{V_1,s\_HDR} = dD_{V_1,HDR}/dD_{V_1}$$
$$= \left(b_0 t_0^2 \frac{D_{Z_0}}{D_{Z_1}} + b_1 t_1^2\right) \left(a_1 + a_2 \left(L_{1,M} + \hat{L}_{1,M}\right)\right)^2 \tag{31}$$

In a general case in which N views are to be encoded, before V_m is encoded, where $m > 0$, m-1 views (i.e., V_0, $V_1, \ldots V_{m-1}$) are already encoded. The HDR distortion for current view V_m is expressed as follows:

$$D_{V_m,HDR} = \left(\sum_{i=0}^{m-1} b_i t_i^2 DR_{i\to m} + b_m t_m^2\right) D_{Z_m}$$
$$= \left(\sum_{i=0}^{m-1} b_i t_i^2 DR_{i\to m} + b_m t_m^2\right)$$
$$\times \left(a_1 + a_2 \left(L_{m,M} + \hat{L}_{m,M}\right)\right)^2 D_{V_m} \tag{32}$$

where $DR_{i\to m}$ is the ratio between D_{Z_i} and D_{Z_m}. $L_{m,M}$ and $\hat{L}_{m,M}$ are the mean values of L_m and $\hat{L}_m$ in the CTU of V_m. The optimal value for λ_{V_m,s_HDR} to encode V_m is derived as follows:

$$\lambda_{V_m,s\_HDR} = dD_{V_m,HDR}/dD_{V_m}$$
$$= \left(\sum_{i=0}^{m-1} b_i t_i^2 DR_{i\to m} + b_m t_m^2\right) \left(a_1 + a_2 \left(L_{m,M} + \hat{L}_{m,M}\right)\right)^2 \tag{33}$$

Using ICRF, Z is estimated using the technique described in Section III-B-2). Parameters a_0, a_1, and a_2 in (25) can be computed coding unit (CU)-wise through least squares fitting.

The original Lagrange multiplier used in HEVC can be expressed as a function of the QP value. The adaptive QP technique can be optionally used in MV-HEVC to increase coding efficiency. In this technique, the QP that provides the optimal RD performance is identified from several QPs around a given QP. In the proposed scheme, the mode for a nonbase view is determined on the basis of the proposed RDO process, in which the distortion and Lagrange multiplier are modified. The adaptive QP technique is adopted in the proposed scheme. For a given CTU, the QP determined in the proposed scheme may differ from that of the original MV-HEVC because of the change in the RD cost. For instance, a CTU may be assigned a low QP in the original MV-HEVC after the original RDO process is performed; however, this CTU may be assigned a high QP in the proposed scheme after the computation of the new RDO cost that indicates the low contribution of the CTU to the generated HDR image. The ΔQP value used in the proposed scheme is 3; thus, seven QPs are examined to determine the optimal QP and corresponding coding mode.

The proposed MV-HEVC based coding scheme can be further extended for the multi-exposure video coding where the input video is captured with alternating exposures. Before encoding, the multi-exposure video is divided into several views where the LDR frames with the same exposure are collected and formed a view. The multi-exposure image coding can be treated as a special case in which the frame number in each view is one.

TABLE 2. Information Regarding Test Exposure Stacks

Sequence	Resolution	Exposure time (s)
Cave	512×384	1/5, 2, 30
Recorder	1024×768	4, 4/5, 1/60
Balloons	896×576	1/60, 1/250, 1/1000
Mask	1024×768	1/45, 1/12, 1/3
Belgium House	512×384	1/8, 1/30, 1/125
Swiss	896×576	1/2, 1/8, 1/30
Estate	864×488	1/120, 1/287, 1/713
Lighthouse	512×336	1/40, 1/10, 1/2

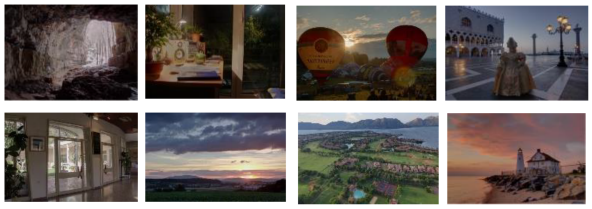


FIGURE 3. MEF images for the test multi-exposure images. Top row (from left to right): Cave, Recorder, Balloons and Mask. Bottom row (from left to right): Belgium House, Swiss, Estate, and Lighthouse.

IV. RESULTS AND DISCUSSION

Multi-exposure images were encoded using MV-HEVC, where the base view (i.e., the well-exposed image) was encoded without any change and two modification steps were performed for the nonbase views: (1) the intensity mapping of the reference view before the encoding of a nonbase view and (2) the modification of the distortion and Lagrange multiplier for the RDO process. For successful decoding, look-up tables were used to store the IMF and ICRF. These functions are entropy encoded, and their bit consumption is low.

To evaluate coding performance, eight multi-exposure images [57] were used. Information on these images is listed in Table 2. The corresponding MEF images generated by the method in [24] are shown in Fig. 3 to visualize the fused images.

A. REFINEMENT OF LAGRANGE MULTIPLIER

The Lagrange multiplier was modeled as a function of QP with a configured scaling, as presented in (7). In the proposed scheme, a scaling factor was adopted to the derived Lagrange multiplier $\lambda_{V,HDR}$ in (20) and final Lagrange multiplier $\lambda^F_{V,HDR}$ is obtained by using the following expression:

$$\lambda^F_{V,HDR} = s \cdot \lambda_{V,HDR} = s \cdot \lambda_{V,s\_HDR} \cdot \lambda, \tag{34}$$

where s is a scaling factor that is empirically determined for the under-exposed image and over-exposed image separately.

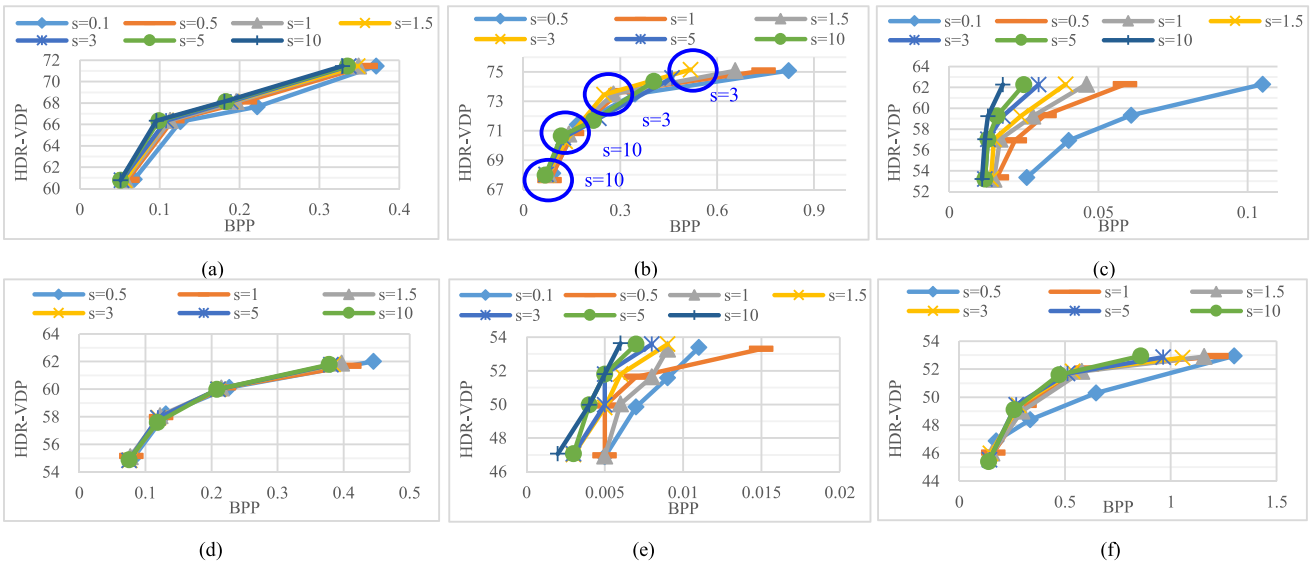


FIGURE 4. RD performance under different scaling factors used in the proposed Lagrange multiplier: (a) under-exposed image of Balloons, (b) over-exposed image of Balloons, (c) under-exposed image of Swiss, (d) over-exposed image of Swiss, (e) under-exposed Recorder image, and (f) over-exposed Recorder image.

Three test images and four QP values {22, 27, 32, 37} were examined and HDR-VDP2 [48] was used to evaluate the HDR image quality, which measures the difference between the reconstructed HDR image and the uncompressed HDR image. A higher HDR-VDP-2 score indicates a higher degree of similarity between two HDR images. The s value of the under-exposed image was determined first because this image was the second view to be encoded.

Fig. 4 depicts the RD performance under different s values. Fig. 4(a), (c), and (e) presents the HDR-VDP2 scores under four QPs when two views, namely the well-exposed image and under-exposed image, were encoded. The HDR image generation and the bitrate consider only the well-exposed and under-exposed images. Fig. 4(a), (c), and (e) indicates that the optimal RD performance was achieved when $s = 10$ for all QPs. The optimal s value for the over-exposed image was similarly determined by examining the RD performance under different s values. After the optimal s value for the under-exposure image was determined, the three exposure images including the well-exposed image, the under-exposed image and the over-exposed image were used in HDR generation. The s value for the under-exposed image was fixed at 10 during the experiments.

Fig. 4(b), (d), and (f) presents the RD performance, and the optimal s value was obtained by determining the highest slope in a given bitrate range. For the Swiss image, a minor difference was observed in the RD performance at different s values. For the Recorder image, the optimal RD performance was achieved when $s = 10$; however, this performance was not significantly superior to that observed for this image at other s values. For the Balloons image, the RD curves varied depending on the s value. In the case of this image, the optimal s value was determined individually for each QP. In this paper, the set of optimal s values was {3, 3, 10, 10} for the QP set of {22, 27, 32, 37} for the over-exposed image.

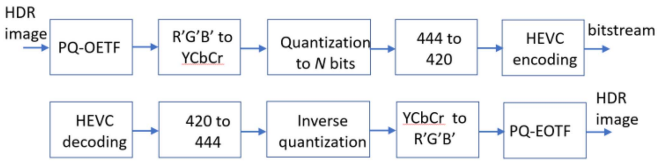


FIGURE 5. Procedures of HDR image encoding-decoding in HEVC (the top row presents encoding and the bottom row presents decoding).

B. RD PERFORMANCE FOR HDR IMAGES

After the final Lagrange multiplier was determined, the RD performance for the test images was evaluated. The performance of several coding schemes was compared with that of the proposed coding scheme. Two of the compared schemes used HDR images as input, whereas the remaining schemes used multi-exposure images as input. HDR images were generated using multi-exposure images and the Robertson ICRF. The compared schemes are as follows:

1) *HEVC-10bit:* HEVC-10bit is a single-layer coding scheme in which a floating-point HDR image is converted into a 10-bit format before HEVC encoding. The scheme is illustrated in Fig. 5 [58].
2) *JPEG-XT:* JPEG-XT is a backward-compatible coding scheme in which Profile C is used. The input is HDR images with the floating-point format
3) *MV-HEVC:* In MV-HEVC, multi-exposure images are encoded without any change.
4) *MV-HEVC (IMF):* In MV-HEVC (IMF), multi-exposure images are encoded using MV-HEVC, and the IMF is applied to the base view during the encoding of the nonbase view.

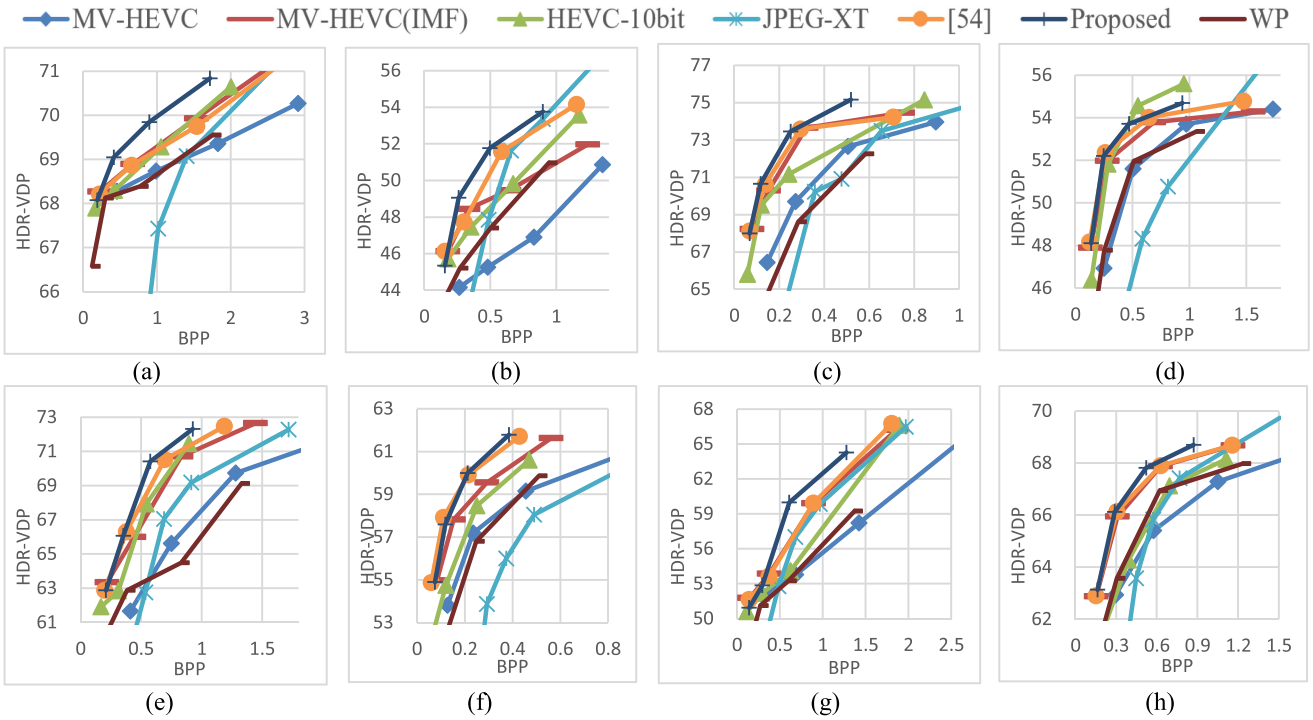


FIGURE 6. RD performance for HDR images when Robertson CRF model was used for HDR generation: (a) cave, (b) recorder, (c) balloons, (d) mask, (e) Belgium house, (f) swiss, (g) estate, and (h) lighthouse.

TABLE 3. Encoding Parameters of Compared Schemes

Standard	MV-HEVC	HEVC	JEPG-XT
Platform	HTM 16.2	HM16.9	
QP/QF	25, 30, 35, 40	7, 12, 17, 22, 25, 27, 32	(QF_base, QF_residual) w/o refinement scan (10, 10), (30, 10), (50, 10), (70, 10), (90, 10) with refinement scan (10, 70), (10, 90)

5) *Method proposed in [54]:* multi-exposure images are encoded using MV-HEVC, and the RDO process is modified by considering also the distortion of MEF the image.
6) *Weighted prediction (WP):* multi-exposure images are arranged as a video and encoded by HEVC with weighted prediction enabled.

The video is composed of a well-exposed image, followed by an under-exposed image and an over-exposed image.Three coding platforms are used in the schemes, and the encoding parameters of these schemes are listed in Table 2. For HEVC-10bit scheme, HEVC Format Range Extension (REext) is used, and the QP for each HDR image is used to produce a bitrate range comparable with the proposed scheme. A refinement scan is used to enhance the reconstruction in JPEG-XT. The quality factors (QFs) of the base and residual layers are determined by optimizing coding efficiency. This study adopted the QF combinations suggested in [54].

Fig. 6 illustrates the RD performance for all the test sequences. The MV-HEVC and JPEG-XT schemes exhibited the worst performance. The poor RD performance of JPEG-XT scheme was attributed to the overhead of two-layer coding. The poor performance of MV-HEVC was attributed to a considerable intensity difference between the views. The MV-HEVC (IMF) scheme resolved this problem by aligning the intensity of the reference view with that of the nonbase view before encoding. Thus, the MV-HEVC (IMF) scheme outperformed the MV-HEVC scheme.

The proposed scheme, the MV-HEVC (IMF) scheme, and the scheme proposed in [54] outperformed the HEVC-10bit scheme, which indicates that coding multi-exposure LDR images is more efficient than coding a single-layer HDR image. After the distortion and Lagrange multiplier were modified in the RDO process, the proposed scheme and the scheme proposed in [54] achieved a higher coding efficiency than did the MV-HEVC(IMF) scheme. In the scheme proposed in [54], the distortion in the MEF image was considered. Although this scheme outperformed several others, low distortion in the MEF image does not always guarantee low HDR distortion. In the proposed scheme, the distortion in the HDR image is considered during the RDO process. Thus, this scheme outperformed the other schemes (Fig. 6).

C. HDR RECONSTRUCTION WITH VARIOUS CRF MODELS

To compute the distortion in the HDR domain, this study adopted the Robertson CRF model. On the decoder side, another CRF model can be used to generate the HDR image

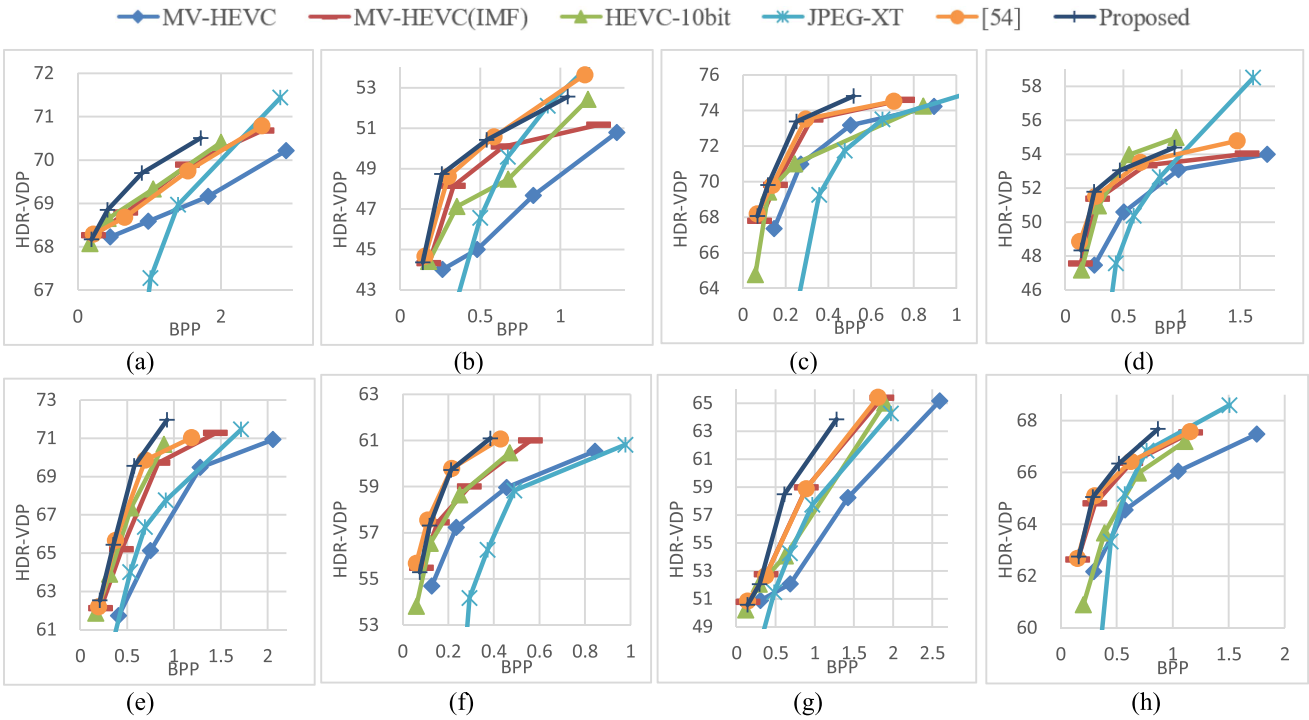


FIGURE 7. RD performance for HDR images when the Devebec CRF model was used for HDR generation: (a) cave, (b) recorder, (c) balloons, (d) mask, (e) Belgium house, (f) swiss, (g) estate, and (h) lighthouse.

after multi-exposure images are reconstructed. To examine the effects of various CRF models on the proposed scheme, the Devebec CRF model [1] is used to generate an HDR image, and the corresponding RD performance in terms of HDR-VDP2 is presented in Fig. 7. The proposed scheme outperformed the other schemes even when the CRF model used at the decoder side differed from that used to derive the HDR distortion and corresponding Lagrange multiplier.

D. RD PERFORMANCE FOR MEF IMAGES

One advantage of the proposed scheme over the single- layer coding scheme, in which an HDR image is generated before encoding, is that the proposed scheme not only can generate HDR image, but also can render MEF images. To fuse multi-exposure images, a pixel-based fusion method [24] was adopted, and Table 4 presents the RD performance in terms of the BD rate [59] for the proposed scheme, where the four anchors were the MV-HEVC scheme, MV-HEVC (IMF) scheme, WP scheme and the scheme proposed in [54]. The proposed scheme outperformed the MV-HEVC, MV-HEVC (IMF) and WP schemes in terms of BD rate. However, on average, the proposed scheme required a 2.3% higher bitrate to render MEF images than did the scheme proposed in [54]. The scheme proposed in [54] modified the distortion in the RDO process by accounting for the reconstructed nonbase view and fused image. Thus, it outperformed the other schemes in rendering MEF images.

Subjective experiments are conducted to evaluate the subjective visual quality for the MEF image under different schemes, where 12 non-experts participate and the Double Stimulus Continuous Quality Scale (DSCQS) [60] method is adopted. Fig. 8 shows the MOS for two images "Cave" and "Swiss". Table 4 shows that, compared to the work in [54], the proposed method required a 6.48% lower BD rate for the Cave image and an 8.93% higher BD rate for the Swiss image. The rate-MOS performance shows similar behavior.

TABLE 4. BD Rate (%) of Proposed Scheme With Each Scheme Being Individually Treated as the Anchor (Considering the PSNR of the MEF Images)

	MV-HEVC	MV-HEVC (IMF)	WP	[54]
Cave	-39.75	-1.97	-60.66	-6.48
Recorder	-38.66	-6.92	-48.89	2.85
Balloons	-50.62	-5.92	-61.63	3.12
Mask	-47.76	-9.56	-57.66	8.87
Belgium House	-50.02	-10.18	-58.02	2.13
Swiss	-39.73	-31.13	-63.11	8.93
Estate	-49.42	-0.88	-54.42	0.17
Lighthouse	-52.90	-6.17	-64.09	-1.19
Average	-46.10	-9.09	-57.51	2.30

E. PERFORMANCE ANALYSIS AND DISCUSSION

The difference between the proposed scheme and the MV-HEVC (IMF) scheme is that the distortion and Lagrange multiplier are modified in the RDO process in the proposed scheme, which increases coding efficiency (Fig. 6). To demonstrate the superiority of the proposed scheme over

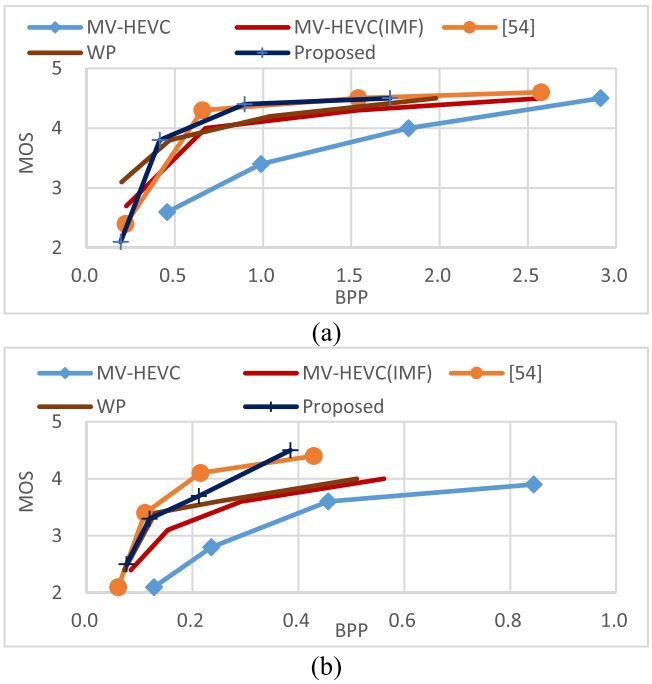


FIGURE 8. Rate-MOS performance for MEF images: (a) cave, (b) swiss.

TABLE 5. CU Partition and Intra/Inter Mode Distribution for Cave Image

QP=25, under-exposed image left: MV-HEVC(IMF), right: proposed								
CU	64×64		32×32		16×16		8×8	
No.	32	41	33	18	69	33	220	28
Intra	11	1	11	3	0	0	4	2
Inter	21	40	22	15	69	33	216	26
DV=0	20	40	19	14	68	33	216	26
QP=25, over-exposed image left: MV-HEVC(IMF), right: proposed								
CU	64×64		32×32		16×16		8×8	
No.	1	30	11	32	246	92	1848	272
Intra	0	0	1	1	77	10	1100	15
Inter	1	30	10	31	169	82	748	257
DV=0	1	30	10	31	158	80	704	250

the MV-HEVC (IMF) scheme, Table 5 presents the coding-parameters, including the coding mode and CU partition, of the schemes for the Cave image under a QP of 25. First, the proposed scheme had more CUs encoded with a size of 64 × 64 than did the MV-HEVC (IMF) scheme; thus, the proposed scheme requires a lower bitrate. Second, the proposed scheme had more CUs encoded in inter mode, particularly for the block size of 64 × 64, than did the MV-HEVC (IMF) scheme, which indicates that inter-view prediction is preferred when distortion from the HDR domain is used. Table 4 lists the number of CUs with zero disparity vectors (DVs). The high percentage of zero DVs indicates that the intensity mapping, which involves changing the intensity of the base view, is accurate and the nonbase view is able to directly take the co-located CU in the reference view for prediction.

The QP map for the Cave image is presented in Fig. 9, and the RD performance for QP = 25 is summarized in Table 6.

TABLE 6. RD Performance for Cave Image With Initial QP of 25

MV-HEVC		MV-HEVC(IMF)		Proposed	
Kbits	PSNR(dB)	Kbits	PSNR(dB)	Kbits	PSNR(dB)
well-exposed image					
178.54	41.73	197.18	42.80	197.18	42.80
under-exposed image					
50.19	45.00	30.45	45.89	1.70	37.11
over-exposed image					
343.98	39.35	279.18	40.58	139.13	36.17
HDR image					
Total bits	HDR-VDP	Total bits	HDR-VDP	Total bits	HDR-VDP
572.71	70.27	506.81	71.16	338.01	70.84

TABLE 7. CU Partition and Intra/Inter Mode Distribution for Lighthouse Image

QP=25, under-exposed image left: MV-HEVC(IMF), right: proposed								
CU	64×64		32×32		16×16		8×8	
No.	18	30	53	36	130	36	168	48
Intra	0	0	2	0	11	2	62	5
Inter	18	32	51	36	119	34	106	43
DV=0	18	30	48	36	115	33	104	43
QP=25, over-exposed image left: MV-HEVC(IMF), right: proposed								
CU	64×64		32×32		16×16		8×8	
No.	24	32	12	19	61	60	716	96
Intra	0	0	0	0	0	0	2	0
Inter	24	32	12	19	61	60	714	96
DV=0	24	32	10	19	61	60	698	94

Fig. 9 indicates that the proposed scheme uses low QP values for CTUs and high Robertson weights (the peak is at an intensity value of approximately 128, as presented in (16)). Fig. 8 shows that the proposed scheme assigns only three CTUs with a QP value of 22, as indicated by a red circle, and these CTUs have high weights. Thus, CTUs that contribute more to a high HDR quality are encoded with less distortion. The MV-HEVC (IMF) scheme contains several CTUs that are encoded with low QPs but with low Robertson weights.

The data in Table 6 indicate that the bit rates required for the encoding of the under-exposed image are 30.45 and 1.7 kbits for the MV-HEVC (IMF) scheme and proposed scheme, respectively. Similarly, the bits required for the encoding of the overexposed image are 279.18 and 139.13 kbits for the MV-HEVC (IMF) scheme and proposed scheme, respectively. Although the MV-HEVC (IMF) scheme reconstructs LDR images with a higher quality and considerably higher bitrate than does the proposed scheme, RD performance in terms of the HDR-VDP of the MV-HEVC (IMF) scheme is inferior to that of the proposed scheme. This result suggests that bits may not be efficiently allocated when only LDR image distortion is considered during multi-exposure LDR image coding and the ultimate goal is to generate high-quality HDR images.

The difference in bit rate between the MV-HEVC (IMF) scheme and proposed scheme was small for the Lighthouse image. To explain this result, Table 7 summarizes the coding mode and CU partition for QP = 25, and Table 8 presents the

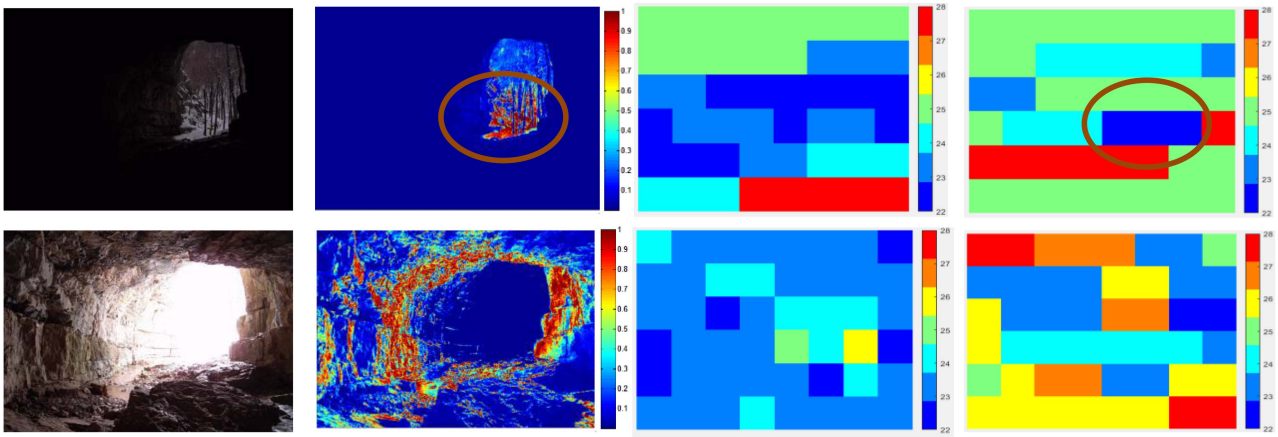


FIGURE 9. Cave images (initial QP = 25). The first row is for the under-exposed image, and the second is for the over-exposed image. The second column presents the Robertson weight map. The third and fourth columns display the QP maps for the MV-HEVC (IMF) scheme and the proposed scheme, respectively.

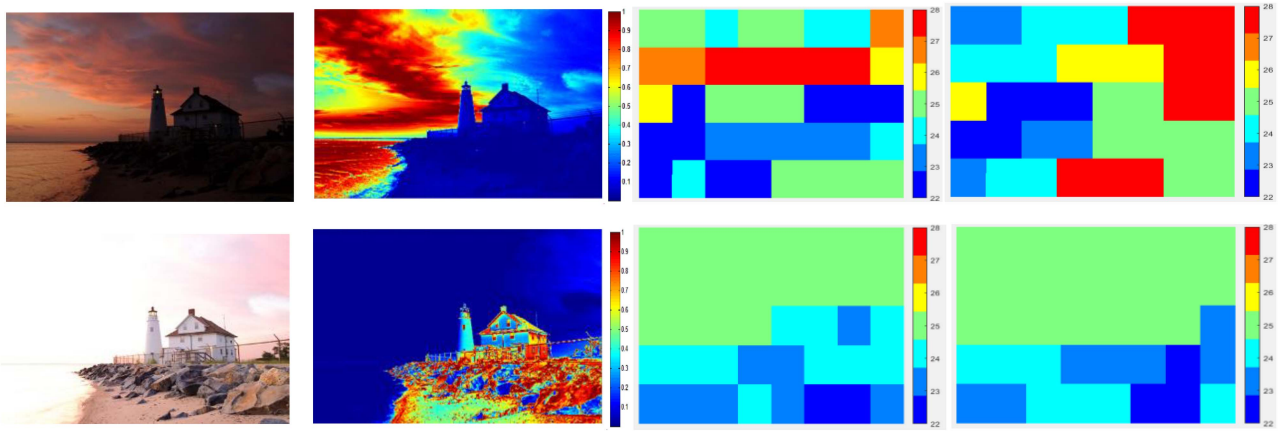


FIGURE 10. Lighthouse images (initial QP = 25). The first row is for the under-exposed image, and the second row is for the over-exposed image. The second column presents the Robertson weight map, and the third and fourth columns present the QP maps for the MV-HEVC (IMF) scheme and the proposed scheme, respectively.

TABLE 8. RD Performance for Lighthouse Image With Initial QP of 25

MV-HEVC		MV-HEVC(IMF)		Proposed	
Kbits	PSNR(dB)	Kbits	PSNR(dB)	Kbits	PSNR(dB)
well-exposed image					
127.88	42.18	144.50	43.22	144.50	43.22
Under-exposed image					
58.65	42.98	24.71	43.62	2.71	37.23
Over-exposed image					
114.86	42.56	30.73	43.94	2.62	39.64
HDR image					
Total bits	HDR-VDP	Total bits	HDR-VDP	Total bits	HDR-VDP
301.39	68.54	199.94	68.68	149.84	68.70

RD performance for QP = 25. The QP maps for the Lighthouse image are presented in Fig. 10. The proposed scheme had more CTUs encoded with a size of 64 × 64 than did the MV-HEVC (IMF) scheme. The QP maps in Fig. 10 indicate that QP has a high correlation with the Robertson weight in the proposed scheme, especially for the under-exposed image. Thus, the bit usage is efficient and only 2.71 kbit in the proposed scheme but 24.71 kbits in the MV-HEVC (IMF) scheme. The bit rates required for the over-exposed image are 30.73 and 2.62 kbits for the MV-HEVC (IMF) scheme and proposed scheme, respectively. This difference in bit rate is considerably smaller than that observed for the Cave image; therefore, the bitrate reduction of the proposed scheme is less remarkable for the Lighthouse image.

V. CONCLUSION

We proposed a multi-view coding scheme to encode multi-exposure images for generating HDR images in the decoder. To ensure high coding efficiency and produce high-quality HDR images, the coding mode involves the selection of a reference frame, block size, and QP. The QP is determined using the proposed RDO process, in which the distortion in the LDR domain is replaced with the distortion in the

HDR domain. To maintain a suitable balance between the rate and the new distortion in the RDO process, a new Lagrange multiplier is derived by modeling the distortion in the HDR domain as a function of the distortion in the LDR domain. The experimental results indicate that the proposed scheme outperformed the other schemes in terms of HDR-VDP-2. The experimental results also demonstrate that the modified RDO process enables a superior HDR image to be obtained given the same bit rate. The proposed scheme outperformed the single-layer coding scheme. The conventional 8-bit codec is still practical for handling HDR image coding if multi-exposure images are used to render an HDR image. In the future, we will extend this work to encode the multi-exposure video coding and generate high-quality HDR video.

REFERENCES

[1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH*, 1997, pp. 369–378.

[2] E. Reinhard, S. Pattanaik, G. Ward, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2006.

[3] S. Battiato, A. Castorina, and M. Mancuso, "High dynamic range imaging for digital still camera: An overview," *J. Electron. Imag.*, vol. 12, no. 3, pp. 459–469, Jul. 2003.

[4] M. D. Grossberg and S. K. Nayar, "Determine the camera response from images: What is knowable?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1455–1467, Nov. 2003.

[5] G. Ward, "Real pixels," in *Graphic Gems II*, J. Arvo, Ed., New York, NY, USA: Academic, 1991, pp. 80–83.

[6] G. Ward, "The LogLuv encoding for full gamut, high dynamic range images," *J. Graph. Tools*, vol. 3, no. 1, pp. 15–31, 1998.

[7] F. Kainz, R. Bogart, and D. Hess, "The OpenEXR image file format," in *Proc. SIGGRAPH Tech. Sketches*, 2003.

[8] T. Richter, A. Artusi, and T. Ebrahimi, "JPEG XT: A new family of JPEG backward-compatible standards," *IEEE MultiMedia Mag.*, Jul./Sep. 2016, doi: 10.1109/MMUL.2016.49.

[9] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 733–741, 2004.

[10] S. Yu, C. Jung, and P. Ke, "Adaptive PQ: Adaptive perceptual quantizer for HEVC main 10 profile-based HDR video coding," in *Proc. IEEE Vis. Commun. Image Process* ., 2016, pp. 1–4.

[11] Y. Liu, N. Sidaty, W. Hamidouche, O. Déforges, G. Valenzise, and E. Zerman, "An adaptive perceptual quantization method for HDR video coding," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 1027–1031.

[12] Y.-J. Chen, W.-H. Shih, J.-C. Chiang, and W.-N. Lie, "Perception-based high dynamic range infrared video coding," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process* ., 2018, pp. 1–5.

[13] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, pp. 267–276, Jul. 2002.

[14] C. Lee and C.-S. Kim, "Gradient domain tone mapping of high dynamic range videos," in *Proc. IEEE Int. Conf. Image Process.*, 2007, pp. 461–464.

[15] J.-C. Chiang and C.-H. Pan, "New tone mapping and tone reproduction techniques-application to bit-depth scalable video coding," in *Proc. IEEE 19th Eur. Signal Process. Conf* ., 2011, pp. 2161–2165.

[16] R. Mantiuk, A. Efremov, K. Myszowski, and H. P. Seidel, "Backward compatible high dynamic range MPEG video compression," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 713–723, 2006.

[17] Z. Wei, C. Wen, and Z. Li, "Local inverse tone mapping for scalable high dynamic range image coding," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 28, no. 2, pp. 550–555, Feb. 2018.

[18] J. Mir, D. S. Talagala, H. K. Arachchi, and A. Fernando, "Adaptive residual mapping for an efficient extension layer coding in two-layer HDR video coding," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1394–1398.

[19] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, "Optimizing a tone curve backward-compatible high dynamic range image and video compression," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1558–1571, Jun. 2011.

[20] A. Koz and F. Dufaux, "Methods for improving the tone mapping for backward compatible high dynamic range image and video coding," *Signal Process.: Image Commun.*, vol. 29, no. 2, pp. 274–292, Feb. 2014.

[21] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert, and J. LeTanou, "Rate-distortion optimization of a tone mapping with SDR quality constraint for backward-compatible high dynamic range compression," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1384–1388.

[22] Z. Mai, H. Mansour, P. Nasiopoulos, and R. K. Kreidieh, "Visually favorable tone-mapping with high compression performance in bit-depth scalable video coding," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1503–1518, Nov. 2013.

[23] M. Le Pendu, C. Guillemot, and D. Thoreau, "Template based inter-layer prediction for high dynamic range scalable compression," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 2974–2978.

[24] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Proc. IEEE 15th Pacific Conf. Comput. Graph. Appl.*, 2007, pp. 382–390.

[25] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.,* vol. 20, no. 12, pp. 3634–3646, Dec. 2011.

[26] K. Ma and Z. Wang, "Multi-exposure image fusion: A patch-wise approach," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 1717–1721.

[27] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, Jan. 2012.

[28] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[29] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[30] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[31] B. Bross et al., "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021.

[32] F. Zhang and D. R. Bull, "An adaptive Lagrange multiplier determination method for rate-distortion optimization in hybrid video codecs," in *Proc. IEEE Int. Conf. Image Process* ., 2015, pp. 671–675.

[33] X. Wang, E.-H. Yang, D.-K. He, L. Song, and X. Yu, "Rate distortion optimization: A joint framework and algorithms for random access hierarchical video coding," *IEEE Trans. Image Process.*, vol. 29, pp. 9458–9469, 2020.

[34] C. Yeo, H. L. Tan, and Y. H. Tan, "On rate distortion optimization using SSIM," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1170–1181, Jul. 2013.

[35] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

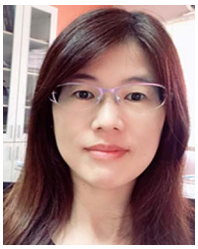
[36] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 806–819, Jun. 2010.

[37] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[38] K. Naser, V. Ricordel, and P. Le Callet, "Modeling the perceptual distortion of dynamic textures and its applocaiton in HEVC," in *Proc. IEEE Int. Conf. Image Process* ., 2016, pp. 3787–3791.

[39] A. Motra and H. Thoma, "An adaptive LogLuv transform for high dynamic range video compression," in *Proc. IEEE Int. Conf. Image Process* ., 2010, pp. 2061–2064.

[40] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High dynamic range video compression exploiting luminance masking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 950–964, May 2016.
[41] J. Mir, G. Kulupana, D. S. Talagala, H. K. Arachchi, and A. Fernando, "Efficient rate-distortion optimization for HDR video compression," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2017, pp. 256–257.
[42] G. Eilertsen, R. K. Mantiuk, and J. Unger, "A high dynamic range video codec optimized by large-scale testing," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1379–1383.
[43] "Image parameter values for high dynamic range television for use in production and international programme exchange," ITU-R BT.2100, 2016.
[44] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usages of bit codes," *SMPTE Motion Imag. J.*, vol. 122, no. 4, pp. 52–59, 2013.
[45] A. Segall, E. François, W. Husak, S. Iwamura, and D. Rusanovskyy, "JVET common test conditions and evaluation procedures for HDR/WCG video," JVET-V2011, Apr. 2020.
[46] S. Choi, O. J. Kwon, J. Lee, and Y. Kim, "A JPEG backward-compatible image coding scheme for high dynamic range images," *Digit. Signal Process.*, vol. 67, pp. 1–16, 2017.
[47] R. K. Mantiuk, T. Richter, and A. Artusi, "Fine-tuning JPEG-XT compression performance using large-scale objective quality testing," in *Proc. IEEE Int. Conf. Image Process* ., 2016, pp. 2152–2156.
[48] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, 2011.
[49] A. Artusi et al., "Overview and evaluation of the JPEG XT HDR image compression standard," *J. Real Time Image Process.*, vol. 16, pp. 413–428, 2019, doi: 10.1007/s11554-015-0547-x.
[50] M. Le Pendu, C. Guillemot, and D. Thoreau, "Inter-layer prediction of color in high dynamic range image scalable compression," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3585–3595, Aug. 2016.
[51] H. E. Egilmez et al., "Transform network architectures for deep learning based end-to-end image/video coding in subsampled color spaces," *IEEE Open J. Signal Process.*, vol. 2, pp. 441–452, 2021.
[52] Y. Wang, D. Liu, S. Ma, F. Wu, and W. Gao, "Ensemble learning-based rate-distortion optimization for end-to-end image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 3, pp. 1193–1207, Mar. 2021.
[53] F. Han, J. Wang, R. Xiong, Q. Zhu, and B. Yin, "HDR image compression with convolutional autoencoder," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process.*, 2020, pp. 25–28,.
[54] J.-C. Chiang, W.-H. Shih, and J.-Y. Deng, "Rate-distortion optimization of multi-exposure image coding for high dynamic range image coding," *Signal Process.: Image Commun.*, vol. 95, 2021, Art. no. 116238.
[55] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vitro, and Y.-K. Wang, Overview of the multiview and 3d extensions of high efficiency video coding, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Sep. 2015.
[56] M. A. Robertson, S. Borman, and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *J. Electron. Imag.*, vol. 36, pp. 162–170, 1999.
[57] Empa HDR Image Database, [Online]. Available: http://www.empamedia.ethz.ch/hdrdatabase/index.php
[58] "Common test condition for HDR/WCG video coding experiments," JCT-VC, Z1020, Jan. 2017.
[59] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," in *Proc. VCEG-M33 Meeting*, Austin, USA, Apr. 2001.
[60] "Methodology for the subjective assessment of the quality of television pictures," *Int. Telecomm. Union/ITU Radio Communication Sector*, ITU-R BT.500-11, Jan. 2002.



JUI-CHIU CHIANG (Senior Member, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1997, the M.S. degree in electronic engineering from National Yang Ming Chiao Tung University, Hsinchu, Taiwan, 1999, and the Ph.D. degree in automatic and signal processing from Paris-Sud University, Orsay, France in 2004. Since August 2005, she has been an Associate Professor with the Department of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan. From August 2012 to January 2013, she was a Visiting Scholar with Signal and Image Processing Department, Télécom Paristech, Paris, France. Her research interests include image/video compression, high dynamic range image/video generation and compression, point cloud compression, learned point cloud compression, and learning-based solutions for computer vision applications.

HUNG-YEN SHANG received the B.S. degree in electrical engineering from National Formosa University, Huwei, Taiwan, in 2019, and the M.S. degree in electronic engineering from National Chung Cheng University, Chia-Yi, Taiwan, 2021. Since September 2021, he has been a Software Engineer with Cubtex. His research interests include signal processing, image/video compression, and high dynamic range image/video compression.

JI-JIN QIU received the B.S. degree in electronic engineering from the National Yunlin University of Science and Technology, Douliu, Taiwan, in 2021. He is currently working toward the M.S. degree with the Department of electronic engineering, National Chung Cheng University, Chia-Yi, Taiwan. His research interests include image/video compression, high dynamic range image/video compression, and learned point cloud compression.